Parameter Plots in Nonlinear Regression

R. Dennis Cook

University of Minnesota
School of Statistics

Technical Report No. 442

August 1984

University of Minnesota
School of Statistics
Department of Applied Statistics
St. Paul, MN 55108

PARAMETER PLOTS IN NONLINEAR REGRESSION

by

R. Dennis Cook
Department of Applied Statistics
University of Minnesota
St. Paul, MN 55108
July 15, 1984

Abstract

We propose graphical methods for displaying the relevant information on a selected parameter from a normal nonlinear regression model. It is shown that the usual extension of added variable plots from linear to nonlinear regression can fail to reveal important diagnostic information and that this information can be recovered by using a parameter plot that depends on selected elements of the parameter-effects curvature array.

Key Words: Added variable plots, compansion, curvature, diagnostics, fanning, graphical methods, parameter-effects curvature, residuals.

1. INTRODUCTION

Diagnostic methods are useful for assessing the adequacy of assumptions underlying the modeling process and for identifying unexpected characteristics of the data that may seriously influence conclusions or require special attention. It is widely held that the diagnostic phase is an important part of any regression analysis.

A variety of diagnostic methods are available to aid in analyses based on linear regression models. For the most part, the development of these methods is based on a thorough characterization of the exact small sample behavior of a few fundamental building blocks such as the ordinary residuals. In more complicated settings such as normal nonlinear regression, the exact small sample behavior of the corresponding building blocks is generally intractable so that some approximation is necessary.

First approximations for nonlinear regression are usually based on the hope that the usual tangent plane is an adequate replacement for the solution locus in a neighborhood of the estimated parameters. While diagnostic methods derived by using this tangent plane approximation are certainly useful and will often provide important information, a somewhat deeper analysis is required for an adequate understanding of nonlinear regression, particularly when substantial curvature is present.

In this paper we report on the extension of added variable plots from linear to nonlinear regression. Detailed background material for linear regression is available from Cook and Weisberg (1982) who use the phrase "added variable plot" to emphasize that these plots are designed to display the information available for assessing the relevance of a selected explanatory variable. In nonlinear regression, however, the idea of an added variable is a bit elusive since there is not necessarily a one-to-one correspondence between parameters and variables. The fundamental objective is to display the essential information available for inference on a selected parameter and thus we use the term parameter plot to refer to the graphical method that is to be developed here.

In section 2 we briefly review added variable plots and show by example that the usual first-order extension of these plots to nonlinear regression can fail to reveal important diagnostic information. In section 3 we use the relevant likelihood confidence region for the selected parameter to develop a parameter plot that recovers the missing information. Section 4 contains examples and our concluding comments are given in Section 5.

2. ADDED VARIABLE PLOTS

2.1 Linear Regression

In the partitioned form of the usual normal linear regression model $\underset{\sim}{Y} = \underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\beta_2 + \underset{\sim}{\varepsilon}$, the standard added variable plot for the single explanatory variable $\underset{\sim}{X}_2$ is obtained by plotting the ordinary residuals $\underset{\sim}{e}_{Y/1}$ from the regression of $\underset{\sim}{Y}$ on $\underset{\sim}{X}_1$ against the ordinary residuals $\underset{\sim}{e}_{2/1}$ from the regression of $\underset{\sim}{X}_2$ on $\underset{\sim}{X}_1$. Such plots are useful for studying the impact of $\underset{\sim}{X}_2$ on the overall regression, obtaining a visual impression of the consistency and strength of relationship, and identifying outliers and influential cases that may seriously impact $\hat{\beta}_2$, the ordinary least squares estimate of β_2.

The plot itself reflects several important features of the overall regression of $\underset{\sim}{Y}$ on $\underset{\sim}{X} = (\underset{\sim}{X}_1, \underset{\sim}{X}_2)$: The slope of the regression through the origin of $\underset{\sim}{e}_{Y/1}$ on $\underset{\sim}{e}_{2/1}$ is $\hat{\beta}_2$, the ordinary residuals from this simple linear regression are the same as those from the overall regression and the usual t-statistic for the hypothesis of a zero slope in the plot is proportional to that for the hypothesis $\beta_2 = 0$ in the overall regression.

Many of the above properties follow immediately from the identity (Cook and Weisberg 1982, p. 46),

$$\underset{\sim}{e}_{Y/1} = \underset{\sim}{e} + \underset{\sim}{e}_{2/1}\hat{\beta}_2 \tag{1}$$

where $\underset{\sim}{e}$ is the vector of ordinary residuals from the regression of $\underset{\sim}{Y}$ on $\underset{\sim}{X}$. The two terms on the right of (1) are orthogonal so that $\underset{\sim}{e}$ determines the scatter in the plot while $\underset{\sim}{e}_{2/1}\hat{\beta}_2$ determines the systematic component. A projection of the plot to have zero slope is equivalent to plotting $\underset{\sim}{e}$ versus $\underset{\sim}{e}_{2/1}$. Clearly, this detrended plot contains essentially the same diagnostic information as the standard form obtained by using (1). It will be convenient to use the detrended version of an added variable plot in the extensions that follow.

2.2 Tangent Plane Extension

The standard nonlinear regression model can be represented as

$$y_i = f(\underset{\sim}{x}_i, \underset{\sim}{\theta}) + \varepsilon_i, \quad i=1,\ldots,n \tag{2}$$

where $\underset{\sim}{x}_i$ represents a vector of known explanatory variables associated with the i-th observable response y_i, $\underset{\sim}{\theta}$ is a p×1 vector of unknown parameters, the response function f is assumed to be twice continuously differentiable in $\underset{\sim}{\theta}$, and the errors are at least tentatively assumed to be independent, identically distributed normal random variables with mean 0 and variance σ^2. Let $\hat{\underset{\sim}{\theta}}$ denote the maximum likelihood estimator of $\underset{\sim}{\theta}$ and for notational convenience let $f_i(\underset{\sim}{\theta}) = f(\underset{\sim}{x}_i, \underset{\sim}{\theta})$.

The first-order extension of the added variable plots described in subsection 2.1 is based on the following development: Let $\underset{\sim}{Y}$, $\underset{\sim}{f}(\underset{\sim}{\theta})$ and $\underset{\sim}{\varepsilon}$ denote n×1 vectors with elements y_i, $f_i(\underset{\sim}{\theta})$ and ε_i, respectively. Rewriting (1) as $\underset{\sim}{Y} = \underset{\sim}{f}(\underset{\sim}{\theta}) + \underset{\sim}{\varepsilon}$, replacing $\underset{\sim}{f}$ with the tangent plane at $\hat{\underset{\sim}{\theta}}$ and rearranging terms gives the linear constructed model

$$\underset{\sim}{e} = \underset{\sim}{V}\underset{\sim}{\phi} + \underset{\sim}{\varepsilon} \tag{3}$$

where $\underset{\sim}{\phi} = \underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}}$, $\underset{\sim}{V}$ is the n×p matrix with elements $\partial f_i/\partial\theta_j$, i=1,...,n, j=1,...,p, and $\underset{\sim}{e} = (e_i)$ is the n×1 vector of ordinary residuals from (2), $e_i = y_i - f_i(\hat{\underset{\sim}{\theta}})$. Here and in what follows all derivatives are evaluated at $\hat{\underset{\sim}{\theta}}$ unless explicitly stated otherwise. Next, partition $\underset{\sim}{\phi}^T = (\underset{\sim}{\phi}_1^T, \phi_2)$ to display the parameter of interest ϕ_2 in the last position and let $\underset{\sim}{V} = (\underset{\sim}{V}_1, \underset{\sim}{V}_2)$ be the conforming partition of $\underset{\sim}{V}$.

Straightforward application of the method described in subsection 2.1 to the partitioned constructed model

$$\underset{\sim}{e} = \underset{\sim}{V}_1\underset{\sim}{\phi}_1 + \underset{\sim}{V}_2\phi_2 + \underset{\sim}{\varepsilon} \tag{4}$$

yields a parameter plot of $\underset{\sim}{e}$ versus $\underset{\sim}{e}_{2/1}$, the ordinary residuals from the ordinary least squares regression of $\underset{\sim}{V}_2$ on $\underset{\sim}{V}_1$. This particular parameter plot for ϕ_2 corresponds to the detrended version of the added variable plot. The original version is obtained simply by replacing $\underset{\sim}{e}$ with $\underset{\sim}{e} + \underset{\sim}{e}_{2/1}\hat{\theta}_2$.

Such parameter plots are certainly useful and will often display relevant diagnostic information. At the same time, however, these plots may fail to reveal critical aspects of the data. To illustrate this potential failure we consider the problem of estimating the coefficient ratio in linear regression through the origin with p=2 explanatory variables. The corresponding nonlinear model can be written as

$$\underset{\sim}{Y} = \theta_1 \underset{\sim}{X}_1 + \theta_1 \theta_2 \underset{\sim}{X}_2 + \underset{\sim}{\varepsilon} \tag{5}$$

For convenience, we assume that $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are orthonormal vectors. Thus, $\hat{\theta}_1 = \underset{\sim}{Y}^T \underset{\sim}{X}_1$, $\hat{\theta}_2 = \underset{\sim}{Y}^T \underset{\sim}{X}_2 / \underset{\sim}{Y}^T \underset{\sim}{X}_1$ and the rotated parameter plot for θ_2 is a plot of $\underset{\sim}{e}$ versus $\underset{\sim}{X}_2 - \hat{\theta}_2 \underset{\sim}{X}_1$. Here and in what follows nonzero scalar multipliers will be ignored when describing the abscissa of parameter plots.

The data in Table 1 were constructed to illustrate the potential failure of the usual parameter plot for θ_2 in model (5). For these data $\hat{\theta}_1 = 100.08$, $\underset{\sim}{Y}^T \underset{\sim}{X}_2 = 2.95$ and $\hat{\theta}_2 = .0295$. A quick inspection of the data in Table 1 shows that the value of X_1 for case 10 is relatively large; this case must be highly influential for $\hat{\theta}_1$ and thus for $\hat{\theta}_2$ also. However, case 10 does not stand out in the plot of $\underset{\sim}{e}$ versus $\underset{\sim}{X}_2 - \hat{\theta}_2 \underset{\sim}{X}_1$ given in Figure 1, so that this parameter plot fails to reveal a relevant aspect of the data. We will return to this example at the beginning of Section 4

3. PARAMETER PLOTS

3.1 Plot Development

To formally extend added variable plots from linear to nonlinear regression we rely on the notion that only $\hat{\beta}_2$, $\underset{\sim}{e}$ and $\underset{\sim}{e}_{2/1}$ are required to set standard confidence intervals for β_2 in linear regression. Collectively, these three quantities can be regarded as a summary of all relevant information on β_2 and thus can be used to display that information.

Let $L(\underset{\sim}{\theta},\sigma^2) = L(\underset{\sim}{\theta}_1,\theta_2,\sigma^2)$ denote the log likelihood for model (2). The standard likelihood region for θ_2 can be represented as (Cox and Hinkley 1974, p. 343)

$$\{\theta_2 | 2[L(\hat{\underset{\sim}{\theta}},\hat{\sigma}^2) - L(\underset{\sim}{m}(\theta_2),\ \theta_2,\ \sigma^2(\theta_2))] \le \rho\} \qquad (6)$$

where ρ, a selected positive constant, is used to set the nominal level, $\hat{\sigma}^2$ is the maximum likelihood estimator of σ^2, and $(\underset{\sim}{m}^T(\theta_2),\ \sigma^2(\theta_2))$ represents the vector-valued function that maximizes L for each fixed value of θ_2. As in linear regression, (6) summarizes essential information on θ_2 and thus may be used to develop graphical displays of that information.

The likelihood region in (6) can be expressed equivalently as

$$\{\theta_2 | n \log[\sum_{i=1}^{n} (y_i - h_i(\theta_2))^2/n\hat{\sigma}^2] \le \rho\} \qquad (7)$$

where $h_i(\theta_2) = f_i(\underset{\sim}{m}(\theta_2),\theta_2)$, $i=1,\ldots,n$. From this representation we see that it will be convenient to work in terms of the constructed model

$$\underset{\sim}{Y} = \underset{\sim}{h}(\theta_2) + \underset{\sim}{\varepsilon} \qquad (8)$$

where $\underset{\sim}{h}$ is the $n\times1$ vector with elements $h_i(\theta_2)$. Using (8) to construct a likelihood region for θ_2 leads back to (7), of course.

The constructed model given in (8) is too complicated to be dealt with exactly, except in certain special cases. We use instead the approximation of (8) obtained by replacing $\underset{\sim}{h}$ with its quadratic expansion about $\hat{\theta}_2$. This yields the constructed model

$$\underset{\sim}{e} = \underset{\sim}{h}_1\phi_2 + \frac{1}{2}\underset{\sim}{h}_2\phi_2^2 + \underset{\sim}{\varepsilon} \tag{9}$$

where $\underset{\sim}{h}_1$ and $\underset{\sim}{h}_2$ are the $n\times 1$ vectors of first and second derivatives of $\underset{\sim}{h}$ with respect to θ_2, respectively. This deceptively simple representation requires considerable further development to extract the relevant information. This development, which is similar to that used by Cook and Goldberg (1984) in their derivation of curvature measures for parameter subsets, seems too detailed for easy reading and thus has been relegated to the Appendix. Here we discuss the final form which is based on the assumption that the intrinsic curvature (Bates and Watts 1980) of $\underset{\sim}{f}$ at $\hat{\underset{\sim}{\theta}}$ is negligible. This assumption is somewhat restrictive but it is valid in the important class of problems where the parameters of interest are nonlinear functions of the location parameters in a linear model. In any event, there seems to be little practical advantage to allowing for a nonnegligible intrinsic curvature since experience has shown (Bates and Watts 1980, Ratkowsky 1983) that it is typically small in practice. Moreover, several additional complications arise when the intrinsic curvature is nonnegligible. For example, the required calculations become much more involved and, as shown by Cook and Tsai (1983), the ordinary residuals cannot be guaranteed to be appropriate diagnostic statistics. The maximum intrinsic curvature can, of course, be evaluated in practice so that this assumption can be checked.

The constructed model (9) is most informatively expressed in terms of the transformed coordinates used by Bates and Watts (1980). Let $\underset{\sim}{V} = \underset{\sim}{U}\underset{\sim}{R}$ denote the unique QR-factorization of $\underset{\sim}{V}$ where $\underset{\sim}{R}$ is upper triangular with positive diagonal elements and the columns of the n×p matrix $\underset{\sim}{U}$ form an orthonormal basis for the column space of $\underset{\sim}{V}$. Next, partition $\underset{\sim}{R}$ as

$$\underset{\sim}{R} = \begin{pmatrix} \underset{\sim}{R}_{11} & \underset{\sim}{R}_{12} \\ \underset{\sim}{0} & R_{22} \end{pmatrix} \tag{10}$$

where $\underset{\sim}{R}_{11}$ is (p-1)×(p-1). The transformed coordinates $\tilde{\underset{\sim}{\phi}}$ can now be defined as $\tilde{\phi}^T = (\tilde{\phi}_1^T, \tilde{\phi}_2) = \underset{\sim}{\phi}^T \underset{\sim}{R}^T$ so that

$$\tilde{\underset{\sim}{\phi}}_1 = \underset{\sim}{R}_{11}\underset{\sim}{\phi}_1 + \underset{\sim}{R}_{12}\phi_2 \tag{11}$$

and

$$\tilde{\phi}_2 = R_{22}\phi_2 \tag{12}$$

Since $\tilde{\phi}_2$ is simply a multiple of $\phi_2 = (\theta_2 - \hat{\theta}_2)$ this transformation will not lose or obscure any information on θ_2.

Let $\underset{\sim}{A}_2$ denote the last p×p face of the p×p×p unscaled parameter-effects curvature array $\underset{\sim}{A}$ as defined in Bates and Watts (1980). In the notation of Bates and Watts, $\underset{\sim}{A} = \underset{\sim}{A}^t/s\sqrt{p}$. The p×p matrix $\underset{\sim}{A}_2$ corresponds to the coordinate of interest $\tilde{\phi}_2$ and is the only face of $\underset{\sim}{A}$ that is relevant to the present development. Next, partition $\underset{\sim}{A}_2$ as

$$\underset{\sim}{A}_2 = \begin{pmatrix} \underset{\sim}{A}_{211} & \underset{\sim}{A}_{212} \\ \underset{\sim}{A}_{212}^T & A_{222} \end{pmatrix} \tag{13}$$

where $\underset{\sim}{A}_{211}$ is $(p-1)\times(p-1)$.

With the above preliminaries the constructed model given in (9) can now be re-expressed as

$$\underset{\sim}{e} = \underset{\sim}{U}_2(\tilde{\phi}_2 + \frac{1}{2} A_{222}\tilde{\phi}_2^2) - \underset{\sim}{U}_1\underset{\sim}{A}_{212}\tilde{\phi}_2^2 + \underset{\sim}{\varepsilon} \tag{14}$$

where $\underset{\sim}{U}_2$ is the last column of $\underset{\sim}{U} = (\underset{\sim}{U}_1, \underset{\sim}{U}_2)$. This form has several revealing features. First, if A_{222} and $\underset{\sim}{A}_{212}$ are sufficiently small, the quadratic components of (14) can be neglected and we obtain the constructed model $\underset{\sim}{e} = \underset{\sim}{U}_2\tilde{\phi}_2 + \underset{\sim}{\varepsilon}$. This model suggests a parameter plot of $\underset{\sim}{e}$ versus $\underset{\sim}{U}_2$. However, $\underset{\sim}{U}_2 = R_{22}^{-1}\underset{\sim}{e}_{2/1}$ so that, apart from an unimportant rescaling of the abscissa, we are led back to the tangent plane extension and constructed model described in subsection 2.2.

Second, if $\underset{\sim}{A}_{212}$ is negligible but A_{222} is not, the constructed model becomes $\underset{\sim}{e} = \underset{\sim}{U}_2(\tilde{\phi}_2 + 1/2\ A_{222}\tilde{\phi}_2^2) + \underset{\sim}{\varepsilon}$. This model does not lead to a new parameter plot, but it does indicate that our interpretation of the plot of $\underset{\sim}{e}$ versus $\underset{\sim}{U}_2$ should include a recognition of A_{222}. Bates and Watts (1981) call A_{222} a compansion term since it reflects the nonlinearity of $\underset{\sim}{f}$ that is due to compression or expansion of scale along the $\tilde{\phi}_2$ parameter curve. The implication of this is that isolated points or unusual configurations in a plot of $\underset{\sim}{e}$ versus $\underset{\sim}{U}_2$ may be contributing to a substantial compansion effect.

Finally, since $\underset{\sim}{U}_1$ and $\underset{\sim}{U}_2$ are orthogonal, we consider the second quadratic component of (14) without reference to the term involving $\underset{\sim}{U}_2$. The important additional knowledge provided by this component is that, in contrast to linear regression, the column space of $\underset{\sim}{U}_1$, which is identical to the column space of $\underset{\sim}{V}_1$, reflects relevant information on θ_2 when $\underset{\sim}{A}_{212}$ is not negligible. The precise nature of this information is determined by the relative magnitudes of the elements of $\underset{\sim}{A}_{212}$ through the new parameter plot of $\underset{\sim}{e}$ versus $\underset{\sim}{U}_1\underset{\sim}{A}_{212}$.

The individual elements of $\underset{\sim}{A}_{212}$ describe the fanning of the p-1 parameter curves associated with $\underset{\sim}{\tilde{\phi}}_1$, as discussed in detail by Bates and Watts (1981). An important implication of this is that remote points in the parameter plot of $\underset{\sim}{e}$ versus $\underset{\sim}{U}_1\underset{\sim}{A}_{212}$ may be influencing $\hat{\theta}_2$ through its relationship with the remaining estimates. It may happen, for example, that a case will influence $\hat{\theta}_2$ indirectly through its influence on $\underset{\sim}{\hat{\theta}}_1$. Generally, this is the kind of important interactive information that is not available in the plot of $\underset{\sim}{e}$ versus $\underset{\sim}{U}_2$, but is recovered by the plot of $\underset{\sim}{e}$ versus $\underset{\sim}{U}_1\underset{\sim}{A}_{212}$.

In rough analogy with response surface models, the plot of $\underset{\sim}{e}$ versus $\underset{\sim}{U}_2$ reflects the main effects of ϕ_2 since it depends only on the uniform tangent plane coordinate system and compansion, while the plot of $\underset{\sim}{e}$ versus $\underset{\sim}{U}_1\underset{\sim}{A}_{212}$ reflects a relevant aspect--fanning--of the "interaction" between $\tilde{\phi}_2$ and the remaining p-1 parameters. To emphasize the contrasting roles of these plots we will henceforth refer to them as the <u>main-effects</u> and <u>interaction</u> parameter plots, respectively.

3.2 Notes on Computation

Although the plots described above require only $\underset{\sim}{U}$, $\underset{\sim}{A}_2$ and $\underset{\sim}{e}$, it will be computationally efficient to have the full p×p×p unscaled parameter-effects array $\underset{\sim}{A}$ available. Bates and Watts (1980) describe methods for computing $\underset{\sim}{A}$ and algorithms are available in Ratkowsky (1983) and Bates, Hamilton and Watts (1983).

The developments in Section 3.1 are based on the assumption that the last element of $\underset{\sim}{\phi}$ corresponds to the parameter of interest. This assumption is necessary to maintain the identity of ϕ_2 when transforming to $\tilde{\phi}_2$, as indicated in (12). The first p-1 faces of $\underset{\sim}{A}$ are associated with the linear combinations $\tilde{\underset{\sim}{\phi}}_1$ of the original parameters given in (12). Thus, the last face $\underset{\sim}{A}_2$ of $\underset{\sim}{A}$ is the only face that can be used directly to construct parameter plots. Generally, it will be necessary to construct new values of $\underset{\sim}{U}$ and $\underset{\sim}{A}_2$ for each parameter of interest. We can, of course, always permute the columns of $\underset{\sim}{V}$ so that the last column corresponds to the parameter of interest and begin again. However, computationally more efficient methods are available for constructing parameter plots for several parameters when the full $\underset{\sim}{A}$ array is available for one ordering of the parameters.

Let $\underset{\sim}{\alpha} = \underset{\sim}{P}\underset{\sim}{\phi}$ where $\underset{\sim}{P}$ is a selected p×p permutation matrix. In what follows, the subscript α added to any quantity indicates evaluation in the permuted $\underset{\sim}{\alpha}$ coordinates. Clearly, $\underset{\sim}{V}_\alpha = \underset{\sim}{V}\underset{\sim}{P}^T = \underset{\sim}{U}\underset{\sim}{R}\underset{\sim}{P}^T = \underset{\sim}{U}\underset{\sim}{U}^*\underset{\sim}{R}^*$ where $\underset{\sim}{U}^*\underset{\sim}{R}^*$ is the QR-factorization of $\underset{\sim}{R}\underset{\sim}{P}^T$. Since the QR-factorization if $\underset{\sim}{V}_\alpha$ is unique, $\underset{\sim}{V}_\alpha = \underset{\sim}{U}_\alpha\underset{\sim}{R}_\alpha$ where $\underset{\sim}{U}_\alpha = \underset{\sim}{U}\underset{\sim}{U}^*$ and $\underset{\sim}{R}_\alpha = \underset{\sim}{R}^*$. From this it can be shown that

$$\underset{\sim}{A}_\alpha = [\underset{\sim}{U}^{*T}][\underset{\sim}{U}^{*T}\underset{\sim}{A}\underset{\sim}{U}^*] \qquad (14)$$

Here and in the remainder of this paper brackets [•][•] indicate column (sample space) multiplication as described in Bates and Watts (1980). In particular, if we let $\underset{\sim}{U}_2^*$ denote the last column of $\underset{\sim}{U}^*$ and let $\underset{\sim}{a}_{ij}$ denote the (i,j)-th column of $\underset{\sim}{A}$, i,j=1,...,p, then the last face $\underset{\sim}{A}_{\alpha 2}$ of $\underset{\sim}{A}_\alpha$ can be written as

$$A_{2\alpha} = U^{*T} B U^{*} \tag{15}$$

where the (i,j)-th element of the p×p matrix B is $U_2^{*T} a_{ij}$.

In short, to construct parameter plots for k parameters we need U and A for one particular ordering and the appropriate permutation matrices P_j, $j=1,\ldots,k-1$, that place the remaining k-1 parameters of interest in the last position. With these and the k-1 QR-factorizations $RP_j^T = U_j^* R_j^*$, $j=1,\ldots,k-1$, the necessary U_α and $A_{2\alpha}$ matrices can be constructed as indicated above.

4. ILLUSTRATIONS

4.1 Model (5)

For our first illustration we return to the example introduced near the end of Section 2. Recall that the main-effects plot for θ_2 is e versus $X_2 - \hat{\theta}_2 X_1$. Similarly, it is easily seen that the interaction plot is e versus $X_1 + \hat{\theta}_2 X_2$ which is essentially a plot of e versus the fitted values from (5). The abscissas of the main-effects and interaction plots are orthogonal, of course.

To understand the usefulness of these two plots we need the A_{222} and A_{212} elements of the 2×2×2 parameter effects curvature array:

$$A_{212} = \frac{1}{\hat{\theta}_1 (1+\hat{\theta}_2^2)^{1/2}} \tag{16}$$

and

$$A_{222} = \frac{-2\hat{\theta}_2}{\hat{\theta}_1(1+\hat{\theta}_2^2)^{1/2}} \tag{17}$$

A variety of useful insights can be obtained from (16) and (17) in combination with the structure described above. A particularly interesting situation occurs when A_{222} is small. Suppose, for example, that $\hat{\theta}_2 = 0$ and that the size of $\hat{\theta}_1$ is highly influenced by a case that has a relatively large y-value at a relatively large X_1-value; removal of this case would result in a substantial change in $\hat{\theta}_1$ and thus a substantial change in $\hat{\theta}_2$. In this extreme but revealing situation, the main-effects plot reduces to simply $\underset{\sim}{e}$ versus $\underset{\sim}{X}_2$ so that there will be little chance to identify the case that is influencing $\hat{\theta}_2$. From the relatively simple structure of this problem it is easily seen that the case in question is influencing $\hat{\theta}_2$ through its dependence on $\hat{\theta}_1$, $\hat{\theta}_2 = \underset{\sim}{Y}^T\underset{\sim}{X}_2/\hat{\theta}_1$, and thus we would expect the influential case to stand out in the interaction plot. This will clearly happen since the interaction plot reduces to $\underset{\sim}{e}$ versus $\underset{\sim}{X}_1$.

In retrospect, the numerical illustration of Section 2 is covered by the above discussion. Figure 2 gives the interaction plot for these data. Case 10 is clearly separated on the right, a usual indication of a highly influential point.

The above discussion is framed in terms of θ_2. When θ_1 is of special interest, $\underset{\sim}{A}_2 = \underset{\sim}{0}$ so that only the main-effects plot is relevant. In this case the main-effects plot for θ_1 is equivalent in construction and interpretation to the corresponding added variable plot.

4.2 Partially nonlinear models

For our second illustration we consider plots for the parameter γ in the partially nonlinear model with response function

$$\underset{\sim}{f}(\underset{\sim}{\theta}) = \underset{\sim}{X}\underset{\sim}{\alpha} + \beta g(\gamma) \tag{18}$$

where $\underset{\sim}{\theta}^T = (\underset{\sim}{\alpha}^T, \beta, \gamma)$, β and γ are scalars and $\underset{\sim}{X}$ is a known $n\times(p-2)$ matrix. In models of this form γ is often of special interest. In particular, (18) allows for transformations of a single explanatory variable in linear regression.

For response function (18)

$$\underset{\sim}{V} = (\underset{\sim}{V}_1 | \underset{\sim}{V}_2) = (\underset{\sim}{X}, g \mid \hat{\beta} g') \tag{19}$$

where g' is the $n\times 1$ vector with elements $g_i' = \partial g_i(\gamma)/\partial\gamma$. It follows immediately from (19) that the main-effects plot for γ is simply a plot of $\underset{\sim}{e}$ versus $\underset{\sim}{e}_{2/1}$, the residuals from the regression of $\hat{\beta} g'$ on $(\underset{\sim}{X}, g)$.

To investigate the interaction plot for γ we need to characterize the column space of $\underset{\sim}{U}_1 \underset{\sim}{A}_{212}$. In general, $\underset{\sim}{A} = [\underset{\sim}{U}^T][\tilde{\underset{\sim}{W}}]$ where $\tilde{\underset{\sim}{W}} = \underset{\sim}{R}^{-T}\underset{\sim}{W}\underset{\sim}{R}^{-1}$ and $\underset{\sim}{W}$ is the $n\times p\times p$ array of second derivatives of $\underset{\sim}{f}$ with respect to $\underset{\sim}{\theta}$; the elements of the i-th face $\underset{\sim}{W}_i$, $i=1,2,\ldots,n$, of $\underset{\sim}{W}$ are the second partial derivatives of $f_i(\underset{\sim}{\theta})$ with respect to $\underset{\sim}{\theta}$. It follows that

$$\underset{\sim}{A}_2 = [\underset{\sim}{U}_2^T][\tilde{\underset{\sim}{W}}] \tag{20}$$

The interaction plot for γ in (18) is easily characterized since $\underset{\sim}{W}_i$ has the relatively simple form

$$\underset{\sim}{W}_i = \begin{pmatrix} \underset{\sim}{0} & \underset{\sim}{b}_{p-1} g_i' \\ g_i' \underset{\sim}{b}_{p-1}^T & \hat{\beta} g_i'' \end{pmatrix} \qquad (21)$$

where $\underset{\sim}{b}_{p-1}$ is the (p-1)-st standard basis vector for R^{p-1} and $g_i'' = \partial^2 g_i / \partial\gamma^2$. The following relations can be verified by using (20) and (21),

$$\begin{aligned} \underset{\sim}{U}_1 \underset{\sim}{A}_{212} &= \underset{\sim}{U}_1 \underset{\sim}{R}_{11}^{-T} \underset{\sim}{b}_{p-1} \underset{\sim}{U}_2^T g'/R_{22} \\ &= \underset{\sim}{V}_1 (\underset{\sim}{V}_1^T \underset{\sim}{V}_1)^{-1} \underset{\sim}{b}_{p-1} \underset{\sim}{U}_2^T g'/R_{22} \\ &= \underset{\sim}{V}_1 (\underset{\sim}{V}_1^T \underset{\sim}{V}_1)^{-1} \underset{\sim}{b}_{p-1}/\hat{\beta} \\ &= \underset{\sim}{Q}_x g/\hat{\beta} || \underset{\sim}{Q}_x g ||^2 \end{aligned} \qquad (22)$$

where $\underset{\sim}{Q}_x$ is the projection operator for the null space of $\underset{\sim}{X}$. Thus, the abscissa of the interaction plot for γ is characterized by $\underset{\sim}{Q}_x g$, the residuals from the regression of g on $\underset{\sim}{X}$.

4.3 Isomerization Model

For our final illustration, which is primarily numerical, we use the data and model described by Box and Hill (1974):

$$f_i(\underset{\sim}{\theta}) = \frac{\theta_0\theta_2(x_{i2}-x_{i3}/1.632)}{1 + \theta_1 x_{i1} + \theta_2 x_{i2} + \theta_3 x_{i3}} \qquad (23)$$

for $i=1,\ldots,24$. Parameter, variable and case indices are the same as those given by Box and Hill who describe a weighted analysis based on the linearized version of (23) obtained by using y_i^{-1} as the response.

Figures 3 and 4 give the main-effects and interaction plots, respectively, for θ_0. Cases 22 and 24 are isolated in Figure 3; these cases appear to be supplying substantial information on θ_0. Nothing appears particularly notable in Figure 4. Evidently, the information on θ_0 that is being supplied by the $\underset{\sim}{U}_1$ component of (14) is spread throughout the data. The main-effects plot seems to supply most of the diagnostic information for θ_0.

The main-effects and interaction plots for θ_1 are given in Figures 5 and 6, respectively. In this case nothing appears to be particularly notable in the main-effects plot, although cases 20 and 22 are somewhat separated from the rest. In the interaction plot, however, cases 20 and 24 are well separated to the right, a usual indication of an influential pair.

In addition to the main-effects and interaction plots illustrated above, we have found that plots of $\underset{\sim}{U}_1$ versus $\underset{\sim}{U}_2\underset{\sim}{A}_{212}$ are also often revealing. For example, Figure 7 contains a plot of $\underset{\sim}{U}_1$ versus $\underset{\sim}{U}_2\underset{\sim}{A}_{212}$ for θ_2. It seems likely that cases 20, 22 and 24 will control inferences concerning θ_2 since they correspond to outlying explanatory variables in the corresponding constructed model (14).

5. DISCUSSION

We have found the graphical methods described in the proceeding sections to furnish important diagnostic information on single parameters in normal nonlinear models. Since the elements of the parameter-effects curvature array $\underset{\sim}{A}$ are typically non-negigible (Bates and Watts 1980, Ratkowsky 1983), we expect that the ability to routinely implement these methods along with the central Bates-Watts procedures should prove to be of substantial value.

The ability of the various plots to display the desired diagnostic information depends on the assumption that $\underset{\sim}{h}$ is quadratic over a sufficiently large neighborhood of $\hat{\theta}_2$. This represents a relaxing of the standard assumption that $\underset{\sim}{h}$ is linear. Further methodology may need to be developed if it is found that $\underset{\sim}{h}$ is often not quadratic and that the plots described here miss relevant information in such cases. At the very least, the proposed methodology is an important and practically useful first step beyond standard tangent plane methods.

ACKNOWLEDGEMENTS

This work was undertaken while the author was a visiting professor at the Mathematics Research Center, University of Wisconsin-Madison, and was sponsored in substantial part by the United States Army under Contract No. DAAG29-80-C-0041 and the Monsanto Company. The author would like to thank K. Larntz and C. L. Tsai for comments on an earlier version of this manuscript, and R. St. Laurent for computational assistance.

APPENDIX

Derivation of Equation (14)

Let $\underset{\sim}{k}^T(\theta_2) = (k_i(\theta_2)) = (\underset{\sim}{m}^T(\theta_2), \theta_2)$ where $\underset{\sim}{m}$ is defined following (6). To derive (14) from (9) we first use the chain rule to express (9) in the equivalent form

$$\underset{\sim}{e} = \underset{\sim}{V}\underset{\sim}{K}_1\phi_2 + \frac{1}{2}\underset{\sim}{K}_1^T\underset{\sim}{W}\underset{\sim}{K}_1\phi_2^2 + \frac{1}{2}\underset{\sim}{V}\underset{\sim}{K}_2\phi_2^2 + \underset{\sim}{\varepsilon} \qquad \text{(A.1)}$$

where $\underset{\sim}{W}$ is defined near (20) and $\underset{\sim}{K}_j$ is the p×1 vector with elements $\partial^j k_i/\partial\theta_2^j$, $i=1,2,\dots,n$, $j=1,2$. Multiplication involving three-dimensional arrays is as defined in Bates and Watts (1980) so that $\underset{\sim}{K}_1^T\underset{\sim}{W}\underset{\sim}{K}_1$ is an n×1 vector with elements $\underset{\sim}{K}_1^T\underset{\sim}{W}_i\underset{\sim}{K}_1$, $i=1,2,\dots,n$.

Explicit forms for $\underset{\sim}{K}_1$ and $\underset{\sim}{K}_2$ can be obtained from the identity

$$\left.\frac{\partial L(\underset{\sim}{m}(\theta_2),\theta_2)}{\partial m_j}\right|_{\underset{\sim}{m} = \underset{\sim}{m}(\theta_2)} = 0 \qquad \text{(A.2)}$$

for $j=1,2,\dots,p-1$ and all θ_2. This identity follows from the fact that $\underset{\sim}{m}(\theta_2) = (m_j)$ maximizes the log likelihood $L(\underset{\sim}{\theta}_1,\theta_2)$ for each value of θ_2.

Let $\underset{\sim}{L}^{(1)}$ and $\underset{\sim}{L}^{(2)}$ denote the p×p matrix and p×p×p array of second and third partial derivatives of L with respect to $\underset{\sim}{\theta}$, respectively, and partition $\underset{\sim}{L}^{(1)}$ as

$$\underset{\sim}{L}^{(1)} = \begin{pmatrix} \underset{\sim}{L}_{11} & \underset{\sim}{L}_{12} \\ \underset{\sim}{L}_{21} & L_{22} \end{pmatrix} \qquad \text{(A.3)}$$

where L_{22} is a scalar. Then differentiating both sides of (A.2) with respect to θ_2 and evaluating at $\hat{\theta}_2$ yields

$$\underset{\sim}{K}_1 = \begin{pmatrix} -\underset{\sim}{L}_{11}^{-1}\underset{\sim}{L}_{12} \\ 1 \end{pmatrix} = \begin{pmatrix} -\underset{\sim}{R}_{11}^{-1}\underset{\sim}{R}_{12} \\ 1 \end{pmatrix} \qquad \text{(A.4)}$$

Here the second equality is obtained by ignoring the intrinsic curvature component $\sum e_i\underset{\sim}{W}$ of

$$\underset{\sim}{L}^{(1)} = \Big(\sum_{i=1}^{n} e_i\underset{\sim}{W}_i - \underset{\sim}{V}^T\underset{\sim}{V}\Big)/\sigma^2 \ .$$

By previous assumption the intrinsic curvature of $\underset{\sim}{f}$ at $\hat{\underset{\sim}{\theta}}$ is negligible.

From (A.4), (10) and (12) it follows that the first term on the right of (A.1) is equal to the first term of (14):

$$\underset{\sim}{V}\underset{\sim}{K}_1\phi_2 = \underset{\sim}{U}_2R_{22}\phi_2 = \underset{\sim}{U}_2\tilde{\phi}_2 \qquad \text{(A.5)}$$

Next, the n×p×p array $\underset{\sim}{W}$ in (A.1) can be decomposed into the sum of three arrays with orthogonal columns,

$$\underset{\sim}{W} = [\underset{\sim}{U}_2\underset{\sim}{U}_2^T][\underset{\sim}{W}] + [\underset{\sim}{U}_1\underset{\sim}{U}_1^T][\underset{\sim}{W}] + [\underset{\sim}{I} - \underset{\sim}{U}\underset{\sim}{U}^T][\underset{\sim}{W}] \qquad \text{(A.6)}$$

Using the first term of (A.6) in the second term on the right of (A.1) and (20) we obtain

$$\frac{1}{2}\,\underset{\sim}{K}_1^T[\underset{\sim}{U}_2\underset{\sim}{U}_2^T][\underset{\sim}{W}]\underset{\sim}{K}_1\phi_2^2 = \frac{1}{2}\,\underset{\sim}{U}_2 A_{222} R_2^2 \phi_2^2$$
$$= \frac{1}{2}\,\underset{\sim}{U}_2 A_{222}\tilde{\phi}_2^2 \qquad (A.7)$$

which is the second addend on the right of (14). The third term on the right of (A.6) represents intrinsic curvature of $\underset{\sim}{f}$ at $\hat{\underset{\sim}{\theta}}$ and is thus neglected.

To this point we have expressed (A.1) in the intermediate form

$$\underset{\sim}{e} = \underset{\sim}{U}_2(\tilde{\phi}_2 + \frac{1}{2}\,A_{222}\tilde{\phi}_2^2) + \frac{1}{2}\,\underset{\sim}{K}_1^T[\underset{\sim}{U}_1\underset{\sim}{U}_1^T][\underset{\sim}{W}]\underset{\sim}{K}_1\phi_2^2$$
$$+ \frac{1}{2}\,\underset{\sim}{V}\underset{\sim}{K}_2\phi_2^2 + \underset{\sim}{\varepsilon} \qquad (A.8)$$

For further progress we must have an explicit expression for $\underset{\sim}{K}_2$. Taking second derivatives of (A.2) with respect to θ_2 yields

$$\underset{\sim}{K}_2 = -\begin{pmatrix} \underset{\sim}{L}_{11}^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & 0 \end{pmatrix} (\underset{\sim}{K}_1^T\underset{\sim}{L}^{(2)}\underset{\sim}{K}_1)$$
$$= \begin{pmatrix} (\underset{\sim}{V}_1^T\underset{\sim}{V}_1)^{-1}\sigma^2 & \underset{\sim}{0} \\ \underset{\sim}{0} & 0 \end{pmatrix} (\underset{\sim}{K}_1^T\underset{\sim}{L}^{(2)}\underset{\sim}{K}_1)$$

so that

$$\frac{1}{2}\,\mathbf{V}\mathbf{K}_2\phi_2^2 = \mathbf{M}(\mathbf{K}_1^T\mathbf{L}^{(2)}\mathbf{K}_1)\phi_2^2\sigma^2/2 \tag{A.9}$$

where $\mathbf{M} = (\mathbf{V}_1(\mathbf{V}^T\mathbf{V}_1)^{-1}, \mathbf{0})$. The terms in $\mathbf{L}^{(2)}$ that involve the residuals e_i represent intrinsic curvature and are again neglected. With this (A.9) can be written as

$$\frac{1}{2}\,\mathbf{V}\mathbf{K}_2\phi_2^2 = -\frac{1}{2}\,\mathbf{M}\{\mathbf{K}_1^T[\mathbf{V}^T][\mathbf{W}]\mathbf{K}_1\}\,\phi_2^2$$

$$-\,\mathbf{M}[\mathbf{K}_1^T\mathbf{V}^T][\mathbf{W}]\mathbf{K}_1\phi_2^2 \tag{A.10}$$

The first term of this expression is exactly the negative of the second term on the right of (A.8) so that the constructed model (A.8) now becomes

$$\mathbf{e} = \mathbf{U}_2(\tilde{\phi}_2 + \frac{1}{2}\,A_{222}\tilde{\phi}_2^2) - \mathbf{M}[\mathbf{K}_1^T\mathbf{V}^T][\mathbf{W}]\mathbf{K}_1\phi_2^2 \tag{A.11}$$

Finally, using (A.5), (20) and the definition of $\tilde{\mathbf{W}}$ it can be shown that

$$\mathbf{M}\,[\mathbf{K}_1^T\mathbf{V}^T][\mathbf{W}]\mathbf{K}_1 = \mathbf{U}_1\mathbf{A}_{212}R_{22}^2$$

Thus, (A.11) is equivalent to (14).

REFERENCES

Bates, D.M. and Watts, D.G. (1980), "Relative Curvature Measures of Nonlinearity," Journal of the Royal Statistical Society, Ser. B, 42, 1-25.

Bates, D.M. and Watts, D.G. (1981), "Parameter Transformations for Improved Approximate Confidence Regions in Nonlinear Least Squares," The Annals of Statistics, 9, 1152-1167.

Bates, D.M., Hamilton, D.C. and Watts, D.G. (1983), "Calculation of Intrinsic and Parameter-Effects Curvatures for Nonlinear Regression Models," Communications in Statistics, Part B--Simulation and Computation, 12, 469-477.

Box, G.E.P. and Hill, W.J. (1974), "Correcting Inhomogeneity of Variance with Power Transformation Weighting," Technometrics, 16, 385-389.

Cook, R.D. and Goldberg, M. (1984), "Assessing the Adequacy of Linear Inference Regions for Parameter Subsets in Nonlinear Regression," in preparation.

Cook, R.D. and Tsai, C.L. (1984), "Residuals in Nonlinear Regression," Technical Summary Report No. 2625, Mathematics Research Center, University of Wisconsin-Madison.

Cook, R.D. and Weisberg, S. (1982), Residuals and Influence in Regression, London: Chapman and Hall.

Cox, R.D. and Hinkley, D.V, (1974), Theoretical Statistics, London: Chapman and Hall.

Ratkowsky, D.A. (1983), Nonlinear Regression Modeling, New York: Marcel Dekker, Inc.

Table 1

Constructed Data for Model (5)

Case No.	Y	X_1	X_2
1	-2.63	.04377	-.4793
2	3.43	.05756	-.3836
3	3.90	.01747	-.2875
4	3.83	.03990	-.1919
5	-1.19	.03196	-.0961
6	6.89	.05498	.4784
7	-2.38	.01112	.1915
8	6.09	.03473	.2870
9	0.22	.05131	.3827
10	100.00	.9924	.0008

Figure 1

Standard parameter plot for θ_2 in model (5).

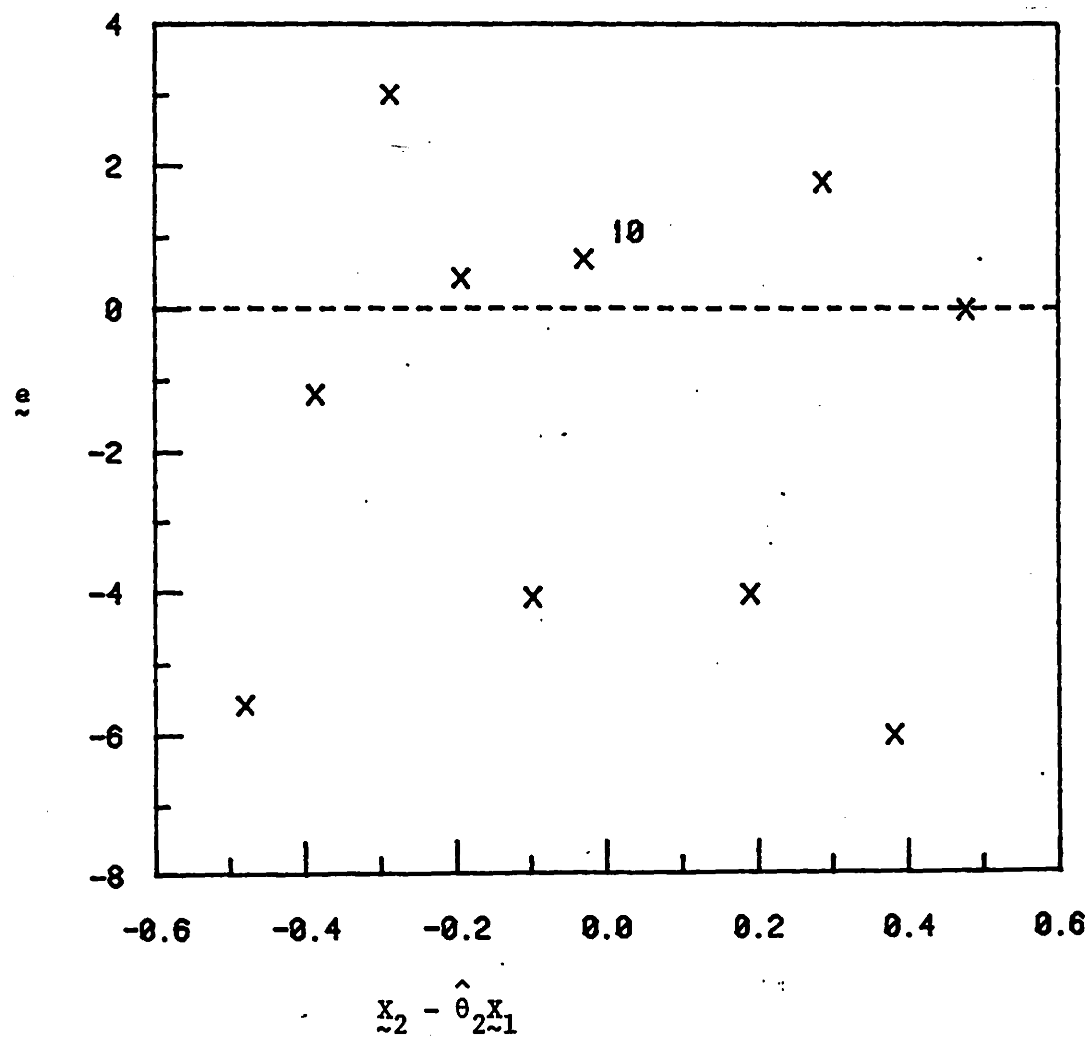


Figure 2

Interaction plot for θ_2 in model (5).

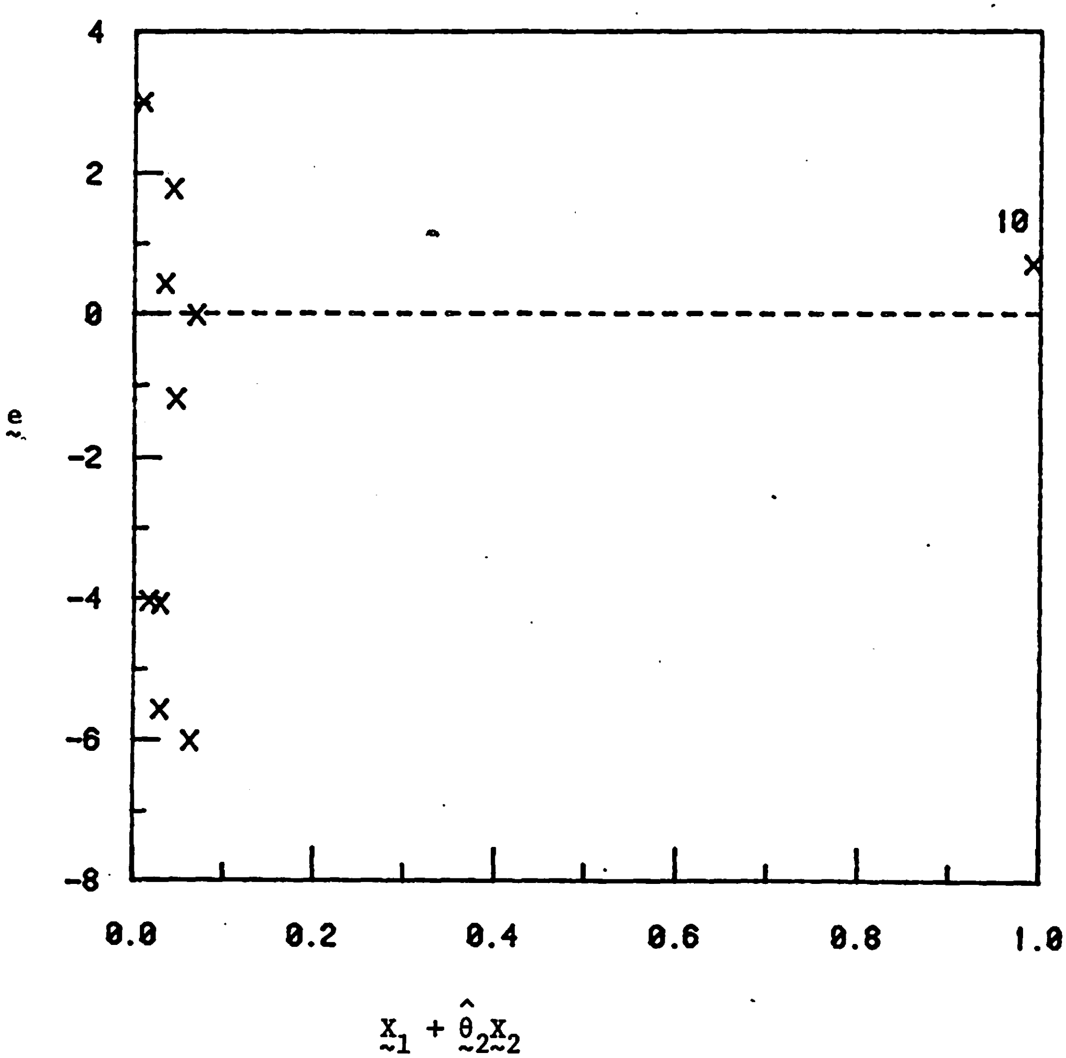


Figure 3

Isomerization model: main-effects plot for θ_0

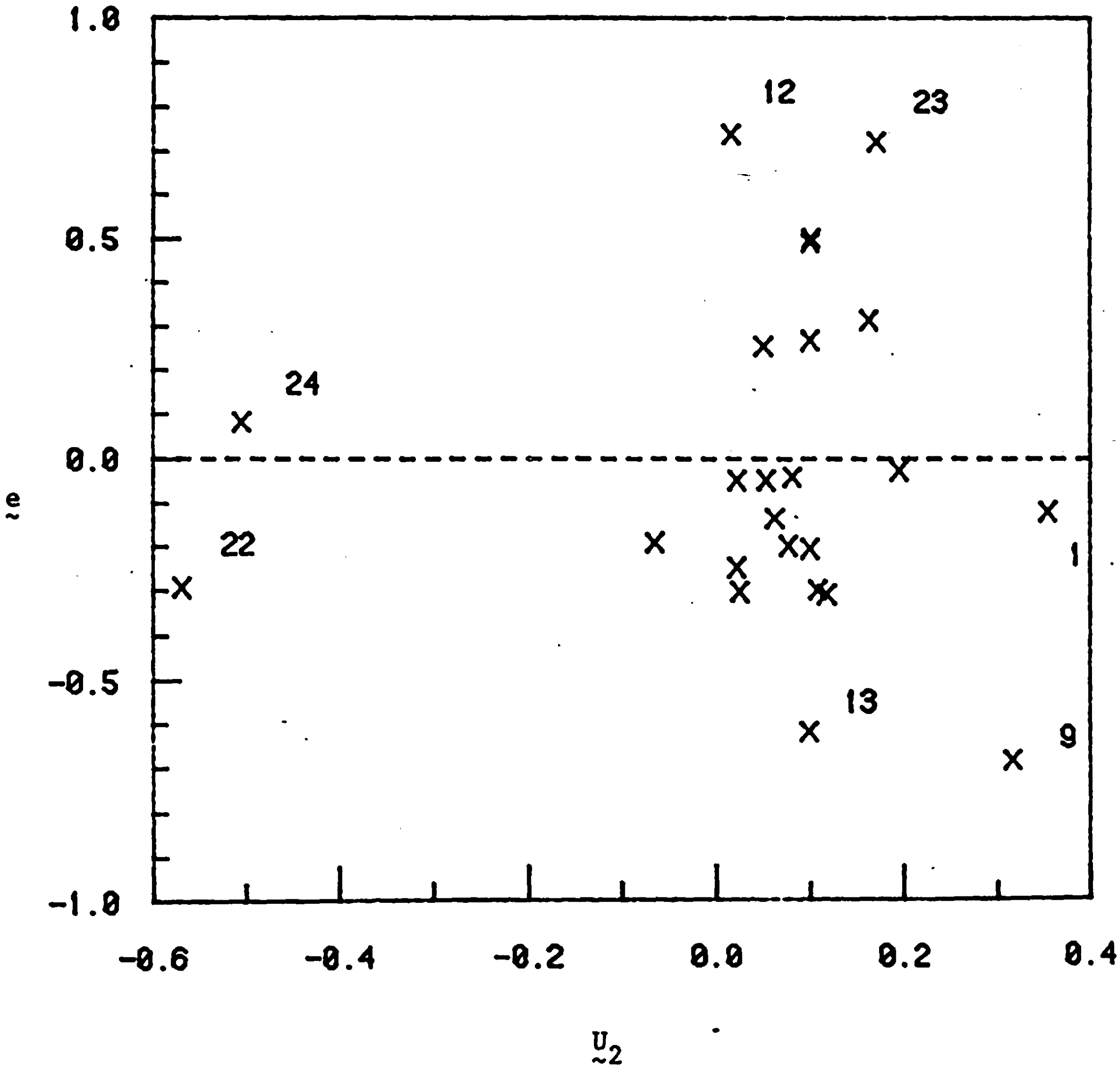


Figure 4

Isomerization model: interaction plot for θ_0

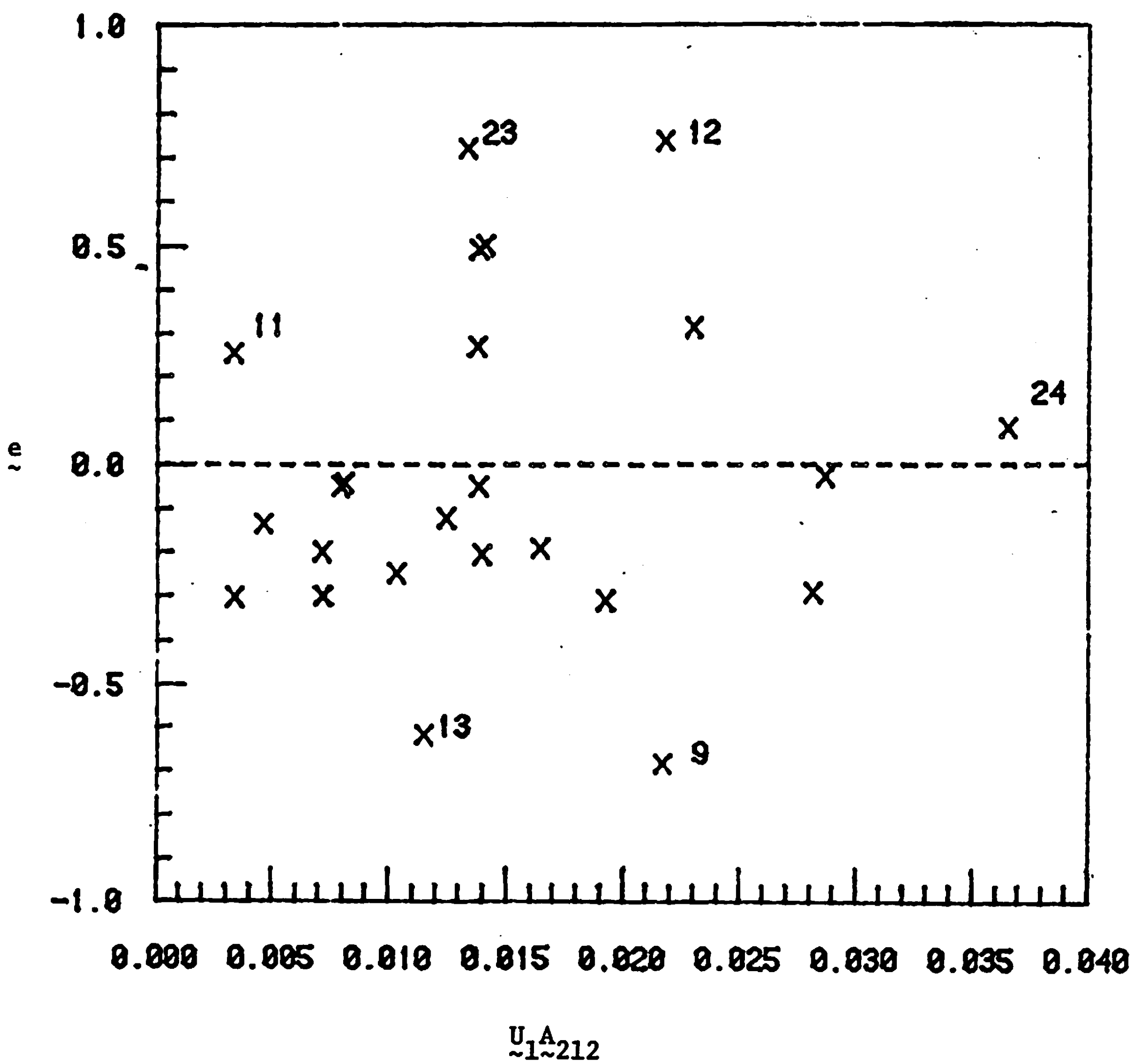


Figure 5

Isomerization model: main-effects plot for θ_1

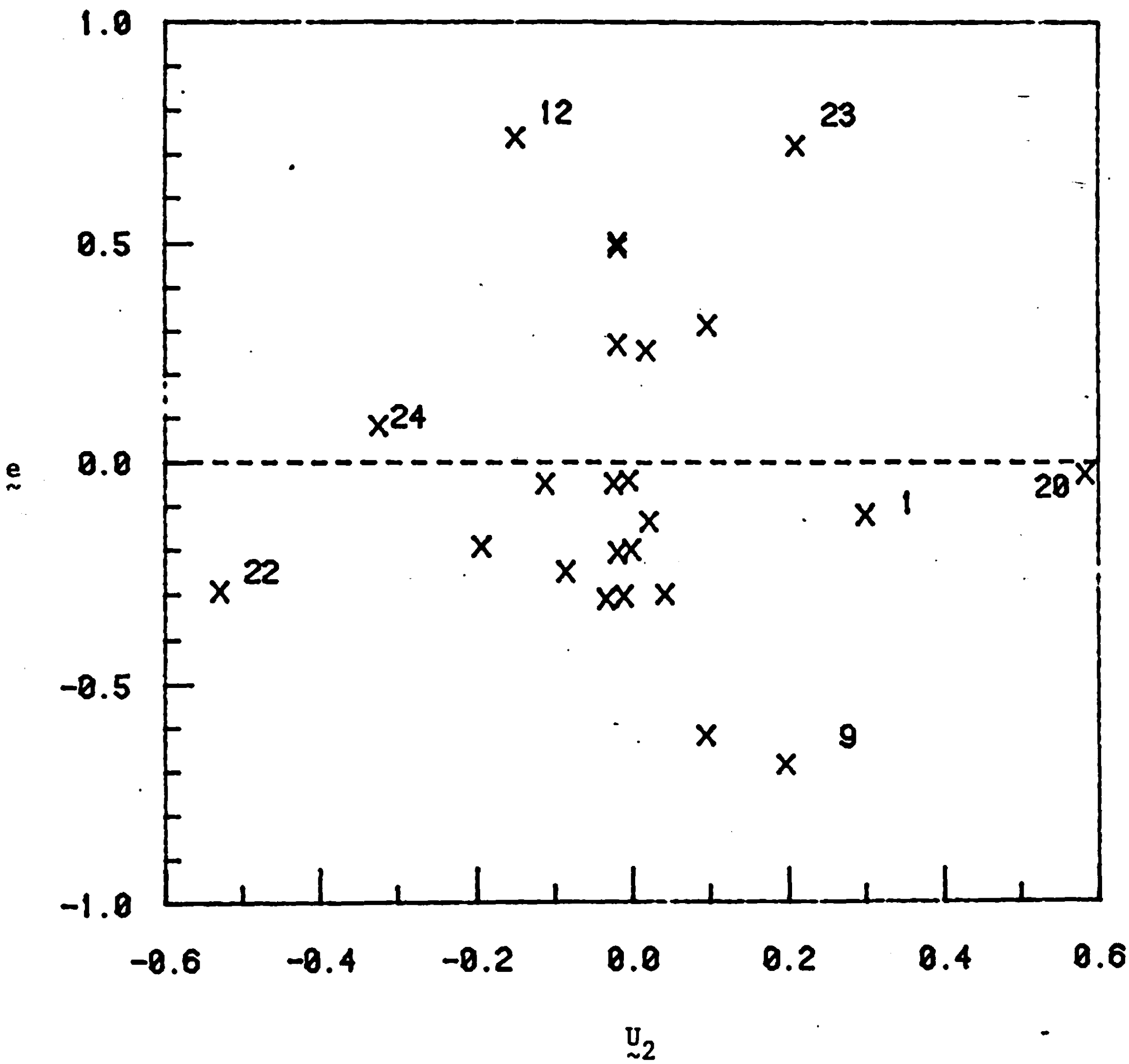
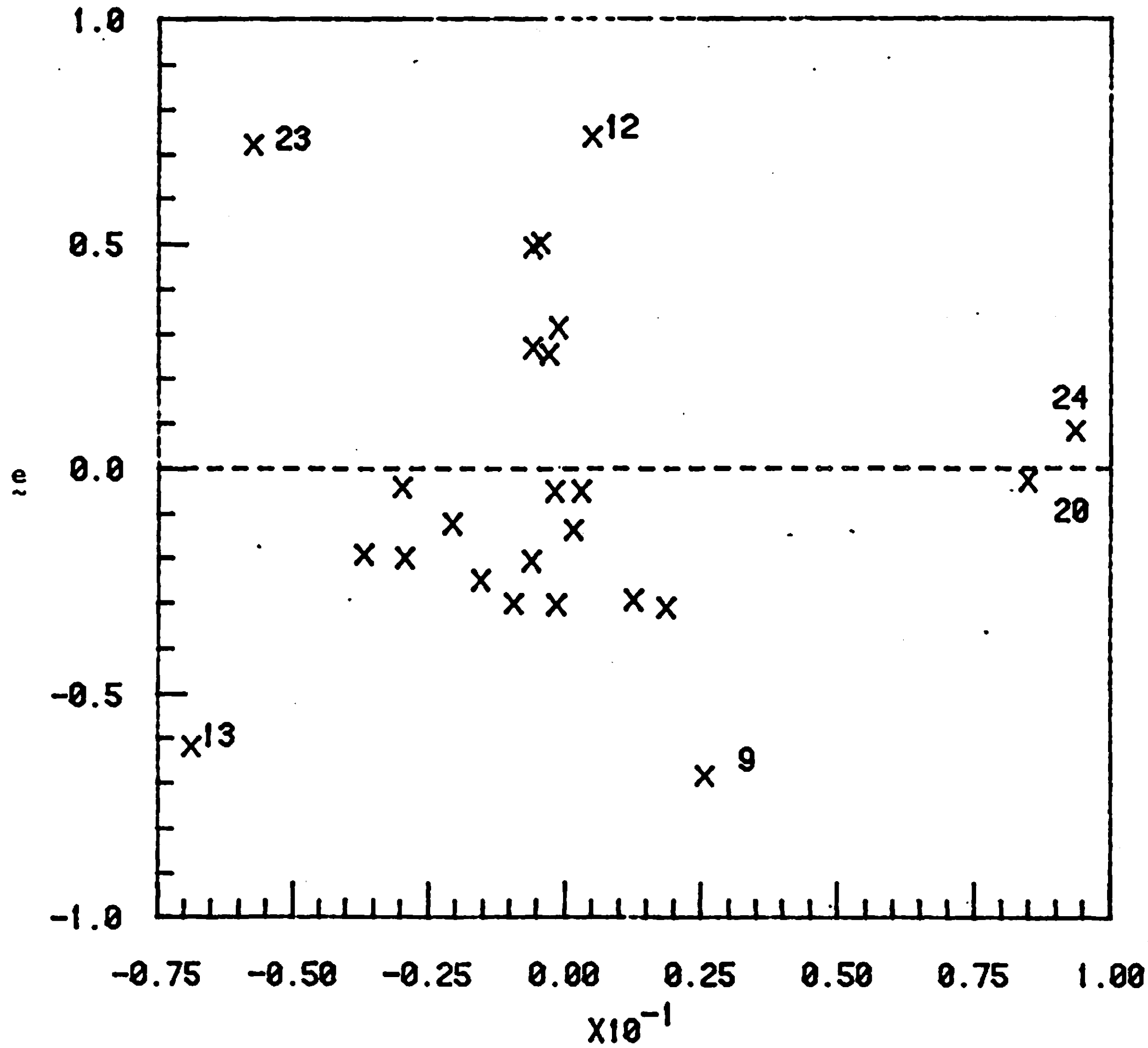


Figure 6

Isomerization model: interaction plot for θ_1



$\underset{\sim}{U}_2\underset{\sim}{A}_{212}$

Figure 7

Isomerization model: scatter plot of

$\underset{\sim}{U}_2$ versus $\underset{\sim}{U}_1\underset{\sim}{A}_{212}$ for θ_2